

On Large Scale Distributed Compression and Dispersive Information Routing for Networks

Kumar Viswanatha*, Sharadh Ramaswamy, Ankur Saxena, Emrah Akyol and Kenneth Rose

{kumar, rsharadh, ankur, eakyol, rose}@ece.ucsb.edu

Department of Electrical and Computer Engineering

University of California at Santa Barbara, CA-93106.

Abstract

This paper considers the problem of distributed source coding for a large network. A major obstacle that poses an existential threat to practical deployment of conventional approaches to distributed coding is the exponential growth of the decoder complexity with the number of sources and the encoding rates. This growth in complexity renders many traditional approaches impractical even for moderately sized networks. In this paper, we propose a new decoding paradigm for large scale distributed compression wherein the decoder complexity is explicitly controlled during the design. Central to our approach is a module called the “bit-subset selector” whose role is to judiciously extract an appropriate subset of the received bits for decoding per individual source. We propose a practical design strategy, based on deterministic annealing (DA) for the joint design of the system components, that enables direct optimization of the decoder complexity-distortion trade-off, and thereby the desired scalability. We also point out the direct connections between the problem of large scale distributed compression and a related problem in sensor networks, namely, dispersive information routing of correlated sources. This allows us to extend the design principles proposed in the context of large scale distributed compression to design efficient routers for minimum cost communication of correlated sources across a network. Experiments on both real and synthetic data-sets provide evidence for substantial gains over conventional approaches.

Index Terms

Large scale distributed compression, dispersive information routing for sensor networks, bit-subset selector

I. INTRODUCTION

The field of distributed source coding (DSC) has gained significant importance in recent years, mainly due to its relevance to numerous applications involving sensor networks. DSC originated in the seventies with the seminal

The work was supported by the NSF under grants CCF-0728986, CCF - 1016861 and CCF-1118075. The results in this paper were presented in part at the IEEE conference on acoustics speech and signal processing (ICASSP) 2010, Dallas, Texas, USA and at the European signal processing conference 2010, Aalborg, Denmark.

work of Slepian and Wolf [1] where they showed that, in the context of lossless coding, side-information available only at the decoder can nevertheless be fully exploited as if it were available to the encoder; in the sense that there is no asymptotic performance loss. Later, Wyner and Ziv [2] derived a lossy coding extension. A number of theoretical publications followed, primarily aimed at solving the general multi-terminal source coding problem (see e.g. [3]). It was not until the late nineties, when first practical DSC schemes were designed adopting principles from coding theory. Today the research in DSC can be categorized into two broad camps. First approach builds on methodologies from channel coding, wherein block encoding techniques are used to exploit correlation [4], [5], [6]. While these techniques are efficient in achieving good rate-distortion performance, they suffer from significant delays and high encoding complexities, which make them unsuitable for many practical applications. The second approach to DSC sprung directly from principles of source coding and quantization [7], [8], [9], [10], [11], [12]. These techniques introduce low to zero delay into the system and typically require minimal encoding complexity. Iterative algorithms for distributed vector quantizer design have been proposed in [8], [9]. A global optimization algorithm, based on deterministic annealing, for efficient design of distributed vector quantizers was proposed in [10]. To optimize the fundamental trade-offs in the (practically unavoidable) case of sources with memory have recently appeared in [11], [12]. Source coding based approaches will be most relevant to us here and will be discussed briefly in Section II.

Distributed coding for a large number of sources is, in theory, a trivial extension of the two source case, but the exponential codebook size growth with the number of sources and encoding rates, makes it nonviable in most practical applications. Just to illustrate, consider 20 sensors, transmitting information at 2 bits per sensor. The base station receives 40 bits using which it estimates the observations of all the sensors. This implies that the decoder has to maintain a codebook of size 20×2^{40} , which requires about 175 Terabytes of memory. In general, for N sources transmitting at R bits, the total decoder codebook size would be $N2^{NR}$. Clearly, the design of optimal low-storage distributed coders is a crucial problem whose solution has existential ramifications to application of DSC in real world sensor networks. Several researchers have addressed this important issue in the past, e.g. [13], [14], [15]. Most of these approaches are based on source grouping mechanisms where the sources are first clustered based on their statistics and then optimal DSC is designed independently within each cluster. However, such approaches suffer from important drawbacks which will be explained in detail in Section II.

In this paper, inspired by our recent results in fusion coding and selective retrieval of correlated sources [16], [17], [18], [19], we propose a new decoding paradigm for large scale distributed coding, where the design explicitly models and controls the decoding complexity. Central to this approach is a new module called the bit-subset selector that allows to judiciously select a subset of the received bits for decoding each source. Specifically, to estimate each source, the bit-subset selector selects only a subset of the received bits that provide the most reliable estimate of

the source. The decoder codebook size then depends only the number of bits selected, which is explicitly controlled during the design. Thus, a direct trade-off between decoder storage complexity and reconstruction distortion is possible. We first present a greedy iterative descent technique for the design of the encoders along with the bit-subset selectors and the decoders and show significant performance improvement over other state-of-the-art approaches. We then present a deterministic annealing (DA) based global optimization algorithm, due to the highly non-convex nature of the cost function, which further enhances the performance over basic greedy iterative descent technique. Experiments with both real and synthetic data sets show that our approach reduces codebook complexity, by factors reaching 16X, over heuristic source grouping methods while achieving the same distortion .

A different problem that is highly relevant to multi-hop sensor networks, and that perhaps surprisingly reveals underlying conceptual similarity with the problem of large scale distributed compression, is that of routing correlated sources across a networks. We recently introduced a new routing paradigm for sensor networks in [20], [21] called “dispersive information routing” (DIR) and showed using information theoretic principles that the new routing technique offers significant (asymptotic) improvements in communication cost compared to conventional routing techniques. In this paper, we point out the close connection between the problem of large scale distributed compression and dispersive information routing and thereby illustrate the general applicability of the approaches herein. We then use similar principles to design efficient low-delay dispersive information routers and demonstrate using both, synthetic and real sensor network grids, that DIR offers significant improvement in performance over conventional routing techniques. We note that a preliminary version of our results appeared in [22] and [23].

The contributions of this paper are summarized as follows:

- 1) We introduce a new decoding paradigm for large scale distributed coding based on a new module at the decoder, called the bit-subset selector, that judiciously selects a subset of the received bits to estimate each source.
- 2) We propose a greedy iterative design strategy for the joint design of the encoders, the bit-subset selectors and the decoders given a training set of source samples and demonstrate using both real and synthetic datasets that the proposed decoding paradigm provides significant improvements in performance over conventional techniques.
- 3) Motivated by the highly non-convex nature of the underlying cost function during design, we propose a deterministic annealing based global optimization approach for the joint design of the system parameters. The proposed design algorithm provides additional gains in performance over the greedy iterative algorithm.
- 4) We then relate the problem of large scale distributed compression to that of multi-hop routing for sensor networks and based on the underlying principles, we propose a new strategy for efficient design of low-delay dispersive information routers. Simulation results on datasets, collected from both real and synthetic sensor

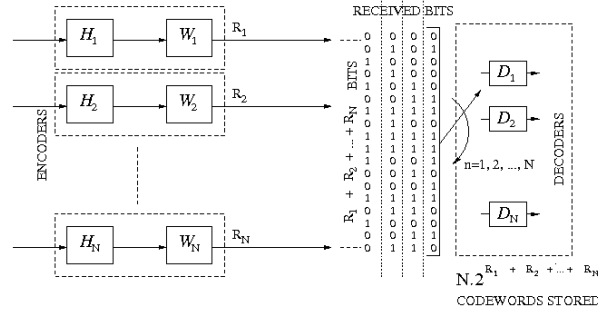


Figure 1. Conventional DSC Setup. Observe that the decoder receives $\sum_i R_i$ bits and has to store a unique reconstruction for every possible received index tuple.

grids, show significant gains over conventional routing techniques.

Much of the focus in the beginning of this paper will be on the problem of large scale distributed compression. In Section II, we begin by describing the canonical distributed source coder and the challenges involved in scaling it to large networks. We then describe the proposed decoding framework for large scale distributed compression based on the bit-subset selector module. In Section IV, we pose the design problem in a constrained optimization framework that allows us to formulate a Lagrangian to quantify the codebook complexity-distortion trade-off. We then propose an efficient algorithm for the joint design of the system components based on deterministic annealing and demonstrate significant gains over heuristic source grouping methods. In Section VI, we show that the proposed methodology has wide applicability beyond the problem of large scale distributed compression, where we extend the principles to design a practical integrated framework for distributed compression and dispersive information routing. We show that the bit-subset selector module and the underlying design methodologies play a central role in designing routers for minimum cost communication of correlated sources across a network.

II. CONVENTIONAL DISTRIBUTED SOURCE CODER

Before describing the problem setup, we state some of the assumptions made in this paper. First, as a simplifying assumption, we consider only spatial correlations and ignore temporal correlations. However, temporal correlations can be easily incorporated using techniques similar to that in [12]. Second, in this paper we assume that the channels from the sensors to the sinks are noiseless/error-free. The design of DSC in the presence of channel noise is an interesting and challenging problem in its own right. Some preliminary research in this directly appeared recently in [24]. In the first half of the paper, we assume that there exists a separate link from every source to the central receiver, i.e., information is not routed in a multi-hop fashion. However, the method we propose is fairly general and is applicable to the multi-hop setting. In the second half of the paper, we focus on DSC design in conjunction with routing in multi-hop sensor networks and show that the methodologies we develop play a central role in efficient

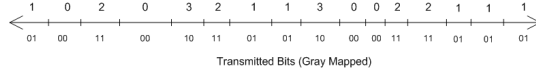


Figure 2. Example of a typical Wyner-Ziv Map

information dispersion across a network. Throughout this paper, we make the practical assumption that while the joint densities may not be known, there will be access to a training sequence of source samples during the design. In practice this could either be gathered off-line before the deployment of the network or could be collected during an initial training phase.

We begin with a description of the conventional DSC system with a single sink. We refer to [10] for a detailed description. Consider N correlated sources, (X_1, X_2, \dots, X_N) transmitting information at rate R_i , respectively, to the sink (fusion center). The sink attempts to jointly reconstruct (X_1, X_2, \dots, X_N) using bits received from all the sources. The setup is shown in Fig. 1 as a block diagram.

The encoding at each source consists of two stages. The first stage is a simple high rate-quantizer which discretizes the input-space into a finite number of regions N_i . Specifically, each high rate quantizer is a mapping:

$$\mathcal{H}_i : \mathcal{R} \rightarrow Q_i = \{1 \dots N_i\} \quad (1)$$

Note that these quantizers are made high rate so as to exclude them from the joint encoder-decoder design. This is a practical engineering necessity, (see e.g., [10]), and the primary purpose of the high rate quantizers is to discretize the source. The second stage, called the ‘Wyner-Ziv map’ (WZ-map), is a module that relabels the N_i quantizer regions with a smaller number, 2^{R_i} , of labels. Mathematically, for each source i , the WZ-map is the function,

$$\mathcal{W}_i : Q_i \rightarrow \mathcal{I}_i = \{1 \dots 2^{R_i}\} \quad (2)$$

and the encoding operation can be expressed as:

$$\mathcal{E}_i(x_i) = \mathcal{W}_i(\mathcal{H}_i(x_i)) \quad \forall i \quad (3)$$

A typical example of a WZ-map is shown in Fig. 2. Observe that the WZ-maps make the encoding operation at each source equivalent to that of an irregular quantizer wherein regions that are far apart are mapped to the same transmission index. Although this operation might seem counter intuitive at first, if designed optimally, it is precisely these modules which assist in exploiting correlations without inter-source communication. The decoder resolves the ambiguity between the regions using the information received from correlated sources. It is fairly well known (see [7]) that these WZ-maps, if properly designed, provide significant improvements in the overall rate-distortion performance compared to that achievable by regular quantizers operating at the same transmission

rates. It is important to note that the WZ-maps must be designed jointly using the training sequence of observations, before the network begins its operation.

The decoder receives $\mathcal{E}_i(x_i) = I_i \forall i$ as shown in the figure. We use $I = \{I_1, I_2, \dots, I_N\}$ to denote the received index tuple and $\mathcal{I} = \mathcal{I}_1 \times \mathcal{I}_2 \dots \mathcal{I}_N$ to denote the set of all possible received symbols at the decoder. The decoder \mathcal{D}_i for each source at the sink is given by a function:

$$\mathcal{D}_i : \mathcal{I} \rightarrow \hat{\mathcal{X}}_i \in \mathcal{R} \quad (4)$$

Usually the decoder is assumed to be a look-up table, which has the reconstruction values stored for each possible received index. For optimal decoding, the look up table has a unique reconstruction stored for each possible received index tuple and hence the total storage at the decoder grows as $\mathcal{O}(N \times 2^{R_r}) = \mathcal{O}(N \times 2^{\sum_{i=1}^N R_i})^1$, which is exponential in N . We refer to the total storage of the look-up table as the decoder complexity. In most prior work, DSC was performed for a few (typically 2 - 3) sources, with the implicit assumption of design scalability with network size. But this exponential growth in optimal decoder complexity with the number of sources and transmission rates makes it infeasible to use the conventional setup in practical settings even with moderately large number of sources. This is one of the major obstacles that has deterred practical deployment of such techniques in real world systems.

One natural solution proposed in the past to handle the exponential growth in decoder complexity is to group the sources based on source statistics [14] and to perform DSC within each cluster. By restricting the number of sources within each group, the decoder complexity is maintained at affordable limits. A major difficulty with such an approach is to come up with good source grouping mechanisms which are optimized for performing distributed compression. While this problem is interesting in its own right, even if optimally solved, such approaches do not exploit inter-cluster dependencies and hence would lead to sub-optimal estimates. We will show in the results section that the proposed bit-subset selector based decoding paradigm provides significant improvements over such source grouping/clustering techniques.

It is worthwhile to mention that an alternate approach, other than the look-up table has been proposed in the literature to practically implement the decoder [13], [15], [25]. In this approach, the decoder computes the reconstructions on the fly by estimating the posterior probabilities for quantization index Q_i as $P(Q_i|I)$, when a particular index tuple is received. Such an approach requires us to store the high rate quantization codewords at the decoder, which grows only linearly in N . However, the computation of the posterior probabilities $P(Q_i|I)$ requires an exponential number of computations, let alone the exponential storage required to store the joint pmf $P(Q_1, Q_2, \dots, Q_N)$. To limit the computational complexity, prior work such as [13], [15], [25] have proposed

¹Note that even for a cascaded coding system, the look-up table at each of the N successive stages of the tree becomes exponentially large.

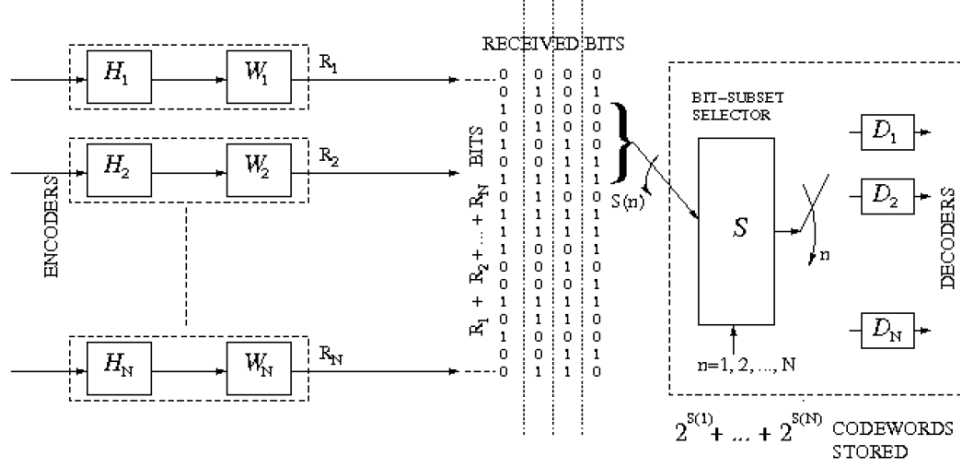


Figure 3. Proposed setup for large scale distributed compression. The bit-subset selector judiciously selects a subset of the received bits for decoding each source. Observe that the decoder now need to only store $\sum_i 2^{S^{(i)}}$ reconstruction codewords.

clustering the sources and linking the clusters using a limited complexity Bayesian network (or a factor graph), and thereby using message passing algorithms to find $P(Q_i|I)$ with affordable complexities. These approaches efficiently exploit inter-cluster dependencies and hence provide significant improvement in distortion over simple grouping techniques for fixed cluster sizes. However, a major shortcoming of these approaches, which is often overlooked, is that the decoder now also needs to store the Bayesian network/factor graph. Though this storage grows linearly in N , it grows exponentially in the ‘rate of the high rate quantizers’. Specifically, if $N_i = 2^{R_q} \forall i$, then the total storage at the decoder grows as $\mathcal{O}(N2^{MR_q})$, where M denotes the maximum number of parents for any source node in the Bayesian network. We demonstrated recently in [24] that this entails a considerable overhead in the total storage required at the decoder and the gains in distortion are often overhauled by the excess storage. We also demonstrated that it is indeed beneficial to group more sources within a cluster instead of connecting the clusters using a Bayesian network. Hence, in this paper, we compare the proposed approach with just the source grouping technique, noting that, the Bayesian network based approaches typically require significantly higher storage to achieve good distortion performance. We refer to [24] for a detailed analysis of the storage requirements for the Bayesian network based techniques.

III. LARGE SCALE DISTRIBUTED COMPRESSION - PROPOSED SETUP

We now describe our approach towards the design of large scale distributed quantizers. Recall that the decoder receives $R_r = \sum_{i=1}^n R_i$ bits of information. For optimal decoding, the decoder needs to store a unique reconstruction in the look up table for every possible received combination of bits. This is infeasible as the decoder complexity grows exponentially in R_r . Hence, we impose a structure on the decoder wherein each source is decoded only

based on a subset of the received bits that provide the most reliable reconstruction for the source. Essentially, we decompose the monolithic decoder, which was a simple look-up table, by introducing an additional module called the ‘bit-subset selector’. For each source, the bit-subset selector determines the subset of the received bits that are used in reconstructing the source. Note that the subset of bits used to estimate each source could be different. Fig. 3 represents our system as a block diagram.

Formally, the bit subset selector is the mapping :

$$\mathcal{S} : \{1, \dots, N\} \rightarrow \mathcal{B} \in 2^{\{1, \dots, R_r\}} \quad (5)$$

where \mathcal{B} is the power set (set of all possible subsets) of the set $\{1, \dots, R_r\}$. For source i , the bit-subset selector retrieves the bits indicated by $\mathcal{S}(i)$ for decoding. This implies that the decoder has to store a unique reconstruction (codeword) for every bit tuple that could be selected. If we denote the decoding rate by $R_{d_i} = |\mathcal{S}(i)|$ bits, the total storage at the decoder is $\sum_{i=1}^N 2^{R_{d_i}}$, where $|\cdot|$ denotes the set cardinality². For different choices of $\mathcal{S}(i)$, reconstructions of source i at different distortion levels are possible and thus, reconstruction quality can be traded against the required codebook size.

The decoder for each source is now modified to be the mapping:

$$\mathcal{D}_i : \mathcal{I} \times \mathcal{B} \rightarrow \mathcal{R} \quad (6)$$

and the reconstruction $\hat{X}_i = \mathcal{D}_i(I, \mathcal{S}(i))$, where $\mathcal{D}_i(\cdot, \cdot)$ depends only on the bits in I at positions indicated by $\mathcal{S}(i)$. We now define the decoder complexity as the average codebook size at the decoder given by,

$$C = \frac{1}{N} \sum_{i=1}^N 2^{R_{d_i}} = \frac{1}{N} \sum_{i=1}^N 2^{|\mathcal{S}(i)|} \quad (7)$$

The average reconstruction distortion is measured as:

$$D = E \left[\sum_{i=1}^N \gamma_i d_i(X_i, \hat{X}_i) \right] \quad (8)$$

where $d_i : \mathcal{R} \times \mathcal{R} \rightarrow [0, \infty) \forall i$ are well defined distortion measures and $0 \leq \gamma_i \leq 1 : \sum_{i=1}^N \gamma_i = 1$ are weights used to measure the relative importance of each of the sources towards the total distortion. Hereafter, we will assume the distortion to be the mean squared error (MSE) and assume equal importance for all the sources, noting that the

²Note that the decoder needs to also store the bit-subset selector module along with the codebooks. However, it is easy to show that the total storage required for the bit-subset selector is $(\sum_i R_{d_i}) \log_2 R_r$ bits and the overhead in storage is typically very small compared to the memory required for the codebook storage.

approach is applicable to more general distortion measures and weights, i.e.,

$$D = \frac{1}{N} E \left[\sum_{i=1}^N (X_i - \hat{X}_i)^2 \right] \quad (9)$$

In practice, we only have access to a training set and not the actual source distributions. Hence, assuming ergodicity, we replace the $E[\cdot]$ operation by a simple empirical average over all the training samples. If the training set is denoted by \mathcal{T} , we measure distortion as:

$$D = \frac{1}{N|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \sum_{i=1}^N (x_i - \hat{x}_i)^2 \quad (10)$$

where $\mathbf{x} = \{x_1, \dots, x_N\}$. We pose the design problem in a constrained optimization framework wherein the objective is to minimize the average distortion (averaged over the training set) subject to a constraint on the total complexity at the decoder. The trade-off between distortion and decoder complexity is controlled by a Lagrange parameter $\lambda \geq 0$ and by optimizing the weighted sum of the two quantities. From (7) and (10), the Lagrangian cost to be minimized is:

$$\begin{aligned} L &= D + \lambda C \\ &= \frac{1}{N|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \sum_{i=1}^N (x_i - \mathcal{D}_i(I, S(i)))^2 \\ &\quad + \frac{\lambda}{N} \sum_{i=1}^N 2^{|S(i)|} \end{aligned} \quad (11)$$

Our objective is to find the optimal encoders (WZ-maps), the bit-subset selectors and the reconstruction codebooks that minimize L for any given value of λ , i.e.,

$$\min_{\mathcal{E}_i, \mathcal{D}_i, \mathcal{S}} L = D + \lambda C \quad (12)$$

IV. ALGORITHMS FOR SYSTEM DESIGN

A natural approach to design such systems is to first derive necessary conditions for optimality of each of the modules and then to iteratively optimize the different modules following a greedy gradient-descent approach. However (12) is a highly non-convex function of the system parameters which makes the greedy approach very likely to get trapped in poor local minima (even when optimized over multiple random initializations), thereby leading to sub-optimal designs. Finding a good initialization for such greedy iterative descent algorithms, even for problems much simpler in nature than the one at hand, is known to be a difficult task. Hence, in this section, we derive a global optimization technique based on deterministic annealing (DA) [26], which has proven to be highly effective in solving related problems in compression and classification. Nevertheless, we begin by describing a greedy-descent algorithm for illustration and ease of understanding. We note that the high-rate quantizers are

designed independently (using a standard quantizer design algorithm) and are excluded from the joint design of the remaining system parameters.

A. Greedy-Descent Algorithm

1) *Necessary Conditions for Optimality:* We first derive the necessary conditions for optimality of all modules in the proposed approach to distributed coding.

1) Optimal Encoders: Let $\mathcal{T}_{i,j} = \{\mathbf{x} \in \mathcal{T} : \mathcal{H}_i(x_i) = j\}$. Then, from (12), the optimal WZ-map, given fixed bit-subset selector and decoder codebooks is:

$$\mathcal{W}_i(j) = \arg \min_{k \in \mathcal{I}_i} \sum_{\mathbf{x} \in \mathcal{T}_{i,j}} \sum_{l=1}^N (x_l - \mathcal{D}_l(I_{i,k}, S(l)))^2 \quad (13)$$

where

$$I_{i,k} = [\mathcal{E}_1(x_1), \dots, \mathcal{E}_{i-1}(x_{i-1}), k, \mathcal{E}_{i+1}(x_{i+1}), \dots, \mathcal{E}_N(x_N)]^T$$

2) Optimal Bit-Subset Selector: For fixed encoders and decoder codebooks, the optimal subset of bits to be used to estimate each source is given by:

$$\mathcal{S}(i) = \arg \min_{e \in \mathcal{B}} \frac{1}{N|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} (x_i - \mathcal{D}_i(I, e))^2 + \frac{\lambda}{N} \times 2^{|e|} \quad (14)$$

3) Optimal Decoders : If $I = [\mathcal{E}_1(x_1), \mathcal{E}_2(x_2) \dots \mathcal{E}_N(x_N)]^T$ represents the bits received from all the sources, and if e represents the positions of bits selected by the bit-subset selector, then we use I_e to denote the index obtained by extracting the bits in I at the positions indicated by e . Differentiating the expression for L , (12), with respect to the reconstruction values and equating it to zero gives the optimal expression for the decoder to be:

$$\hat{x}_i(I_e) = \mathcal{D}_i(I, e) = \frac{1}{|\mathcal{F}|} \sum_{\mathbf{x} \in \mathcal{F}} x_i \quad (15)$$

where $\mathbf{x} \in \mathcal{F}$ if $(\{\mathcal{E}_1(x_1), \mathcal{E}_2(x_2) \dots \mathcal{E}_N(x_N)\})_e = I_e$.

Given the necessary conditions for optimality, a natural design rule is to iteratively optimize the different modules. The algorithm begins with random initializations of all the system parameters. Then, all the parameters are iteratively optimized using (13), (14) and (15) until convergence. When each module is optimized, it leads to a reduction in the Lagrangian cost. Since there are only a finite number of partitions of the training set, convergence to a locally optimal solution is guaranteed. By varying λ , the trade-off between decoder complexity and distortion is controlled and an operational complexity-distortion curve is obtained. The system is optimized over multiple random initializations to avoid poor local minima. However, as we will show in the following section, an algorithm based on DA performs significantly better than such a greedy-descent technique, even with multiple random initializations.

2) *Low Complexity Search:* Before we describe the DA based algorithm for system design, there is a caveat that needs to be addressed. First, let us consider the design complexity of WZ maps. The WZ-maps are updated using (13) which entails a complexity of $\mathcal{O}(N|\mathcal{T}|)$ per source. Hence, the overall complexity of updating WZ-maps in each iteration grows as $\mathcal{O}(N^2|\mathcal{T}|)$, which is quadratic in the number of sources and hence is assumed affordable³. However, the design of the bit-subset selector involves considerably higher complexity. In order to find the best bit-subset selector, every possible configuration of the bit-subset selector must be considered for decoding each source. This will necessitate $\mathcal{O}(N2^{R_r})$ calculations and entails a storage of $N \times \sum_{k=1}^{R_r} \binom{R_r}{k} 2^k = N(3^{R_r} - 1)$ codewords during design. Recall that R_r grows linearly in N , thus making the net complexity cost of updating the bit-subset selector, exponential in N . Hence, for the design algorithm described above, the number of computations and codewords to be maintained grows exponentially in N , which makes it infeasible to implement in practice. Thus we resort to a heuristic approach wherein, instead of finding the best among all possible bit-subset selectors in each iteration of the optimization, we choose an incrementally better solution. Specifically, the search is restricted to only those bit-subsets that differ from the current bit-subset selector setting in exactly one position, i.e., we restrict the search to R_r bit subsets at a Hamming distance of 1 from the current solution. It is easy to verify that the design complexity of such a heuristic technique is $\mathcal{O}(N^2 R_r)$, which is only cubic in N . An added advantage is that this approach requires far smaller storage during design as during each iteration, only R_r codebooks need to be maintained. As we will see in Section V, such a low-complexity design technique performs very close to the full-complexity search for small networks indicating that the loss in optimality due to such a heuristic approach is minimal.

B. Deterministic Annealing Based Design

In this section, motivated by the highly non-convex nature of the cost function in (11), we derive a deterministic annealing (DA) based algorithm for the system design. The approach derives its principles from [26] and builds upon the DA derivation proposed in [10] to encompass the design of the bit-subset selectors jointly with the design of the WZ-maps and the reconstruction codebooks.

A formal derivation of the DA algorithm is based on principles borrowed from information theory and statistical physics. Here, during the design stage, we cast the problem in a probabilistic framework, where the standard deterministic WZ-maps are replaced by a probabilistic mapping, i.e., each of the N_i regions associated with the high rate quantizers are now assigned to each of the 2^{R_i} transmission indices in probability. The expected cost is then minimized subject to an entropy constraint that controls the “randomness” of the solution. By gradually relaxing the entropy constraint we obtain an annealing process that seeks the minimum cost solution. It is important

³Note that, before iterative optimization, the design of the high rate quantizers for all source is performed using the Lloyd-Max scheme, which is a relatively low complexity step.

to note that the WZ-mappings are made soft only during the design stage. Of course, our final objective is to design hard WZ-mappings which minimize the average Lagrangian cost for a fixed λ . More detailed derivation and the principle underlying DA can be found in [26].

In the proposed design approach, we seek to optimize the WZ-maps and the reconstruction codebooks for a fixed bit-subset selector using DA. The bit-subset selectors are then updated to an incrementally better solution using the low complexity update step described in the previous section. This process is repeated until convergence. Hence, during each iteration of the annealing process, the bit-subset selector, and hence the decoder codebook complexity, is fixed. Thus the Lagrangian cost in (12) is determined only by the average distortion, while the codebook complexity term can be temporarily neglected from the cost function. Specifically, during the design, we map the quantization region $q_i \in Q_i = \{1, \dots, N_i\}$ to the transmission index $k_i \in \mathcal{I}_i = \{1, \dots, 2^{R_i}\}$ with probability $P_i(k_i|q_i) \forall i \in \{1, 2, \dots, N\}$. This implies that each of the training set samples are assigned to the transmission indices in probability determined by:

$$P(k_1, k_2, \dots, k_N | x_1, x_2, \dots, x_N) = \prod_{i=1}^N P_i(k_i | x_i) \quad (16)$$

$\forall k_i \in \mathcal{I}_i, \forall (x_1, x_2, \dots, x_N) \in \mathcal{T}$, where

$$P_i(k_i | x_i) = P_i(k_i | \mathcal{H}_i(x_i)) \quad (17)$$

We note that the independence assumption in (16) is required to ensure that the WZ-maps obtained at the end of the annealing process operate independently at the respective encoders. The average distortion is now measured as:

$$\begin{aligned} D_{avg} &= \frac{1}{N|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \sum_{k_1, k_2, \dots, k_N} \left\{ p(k_1, k_2, \dots, k_N | \mathbf{x}) \right. \\ &\quad \left. \times \sum_{i=1}^N (x_i - \mathcal{D}_i(k_1, k_2, \dots, k_N, \mathcal{S}(i)))^2 \right\} \end{aligned} \quad (18)$$

We seek to minimize the above average distortion subject to a constraint on the average randomness in the system. The system randomness is measured using the Shannon's entropy function given by:

$$H = -\frac{1}{N|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \sum_{i=1}^N \sum_{k_i \in \mathcal{I}_i} P(k_i | x_i) \log(P(k_i | x_i)) \quad (19)$$

The DA algorithm minimizes D_{avg} in (18) (for a fixed bit-subset selector), with a constraint on the entropy of the system, (19), where the level of randomness is controlled by a Lagrange parameter (usually called the temperature

of the system due to its roots in statistical physics), T as:

$$J = D_{avg} - TH \quad (20)$$

Initially, when T is set very high, our objective is to maximize the entropy of the system and hence the quantization regions are mapped to all the transmission indices with equal probability. Then during each iteration of the annealing process, the temperature is gradually lowered maintaining J at its minimum. For example, during each iteration, T is updated as $T \rightarrow \alpha T$, where $\alpha = 0.9$. At each temperature, the probabilities $P_i(k_i|q_i)$ and the reconstruction codebooks are perturbed and then iteratively updated until the system reaches equilibrium. Towards deriving the update equations for the probabilities, note that (20) is convex in $P_i(k_i|q_i)$ and hence $\forall k_i \in \{1, \dots, 2^{R_i}\}$, $q_i \in \{1, \dots, N_i\}$, the optimum expression for $P(k_i|q_i)$ is given by the following Gibbs distribution (refer to [10], [26] for the details):

$$P(k_i|q_i) = \frac{e^{-\frac{D_{q_i, k_i}}{T}}}{\sum_{k_i \in \mathcal{I}_i} e^{-\frac{D_{q_i, k_i}}{T}}} \quad (21)$$

where

$$D_{q_i, k_i} = E \left[D \middle| \mathcal{H}_i(x_i) = q_i, p(k_i|q_i) = 1 \right] \quad (22)$$

$$= \sum_{\mathbf{x} \in \mathcal{T} : \mathcal{H}_i(x_i) = q_i} \sum_{k_1, \dots, k_N} \left\{ \prod_{j=1, j \neq i}^N P_j(k_j|x_j) \right. \\ \left. \times \sum_{j=1}^N (x_j - \mathcal{D}_j(k_1, k_2, \dots, k_N, \mathcal{S}(j)))^2 \right\} \quad (23)$$

i.e., D_{q_i, k_i} is the average distortion if the quantization region q_i is mapped to the transmission index k_i deterministically. For fixed $P(k_i|q_i)$, the reconstruction codewords are updated using the following equation:

$$\hat{x}_i(I_e) = \frac{1}{|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} P(I_e|\mathbf{x}) x_i \quad (24)$$

where I_e denotes the subset of the received bits at positions denoted by e .

Hence at each temperature, the reconstruction codebooks and the probabilities are updated iteratively using (21) and (24) till an equilibrium is reached. Finally as $T \rightarrow 0$, $P(k_i|q_i)$ in (21) converge to hard mappings and we obtain the optimum WZ-maps and the codebooks that minimize L for the given bit-subset selector.

Note on Complexity: Observe that, if we use (18) to compute the average distortion, the summation over all k_1, k_2, \dots, k_N requires an exponential number of computations to be performed, i.e., the design complexity of DA grows as $\mathcal{O}(N2^{R_r})$. However, a simple trick that exploits the fact that the bit-subset selector is deterministic while the randomness is all in the WZ maps, allows us to solve this problem. Specifically, the distortion in (18) can be

expressed alternatively as:

$$D = \frac{1}{N|\mathcal{T}|} \sum_{\mathbf{x} \in \mathcal{T}} \sum_{i=1}^N \sum_{s_i \in \mathcal{I}(\mathcal{S}(i))} P(s_i|\mathbf{x}) (\hat{x}_i(s_i) - x_i)^2 \quad (25)$$

where $\mathcal{I}(\mathcal{S}(i))$ denotes the set of all index tuples that can be selected by the bit-subset selector to decode source i and hence the cardinality of $\mathcal{I}(\mathcal{S}(i))$ is $2^{|\mathcal{S}(i)|}$. Observe that the above operation requires far fewer computations, requiring a complexity of $\mathcal{O}(|\mathcal{T}| \sum_{i=1}^N 2^{|\mathcal{S}(i)|})$, which is only linear in N . We further note that computation of D_{q_i, k_i} is effectively an operation that involves finding distortions and hence the complexity of updating the probabilities in (21) grows quadratic in N . Further, the codebook update operation in (24) is a low complexity operation requiring computations only of the order of N . Thus the overall complexity of the DA based algorithm grows of the same order as that for the greedy iterative descent algorithm discussed earlier. We further note that the design is a one-time process and needs to be performed only once before operation. The computational complexity during run-time is negligible and is independent of the algorithm used for the design.

V. RESULTS FOR LARGE SCALE DISTRIBUTED COMPRESSION

We tested our proposed algorithm extensively on both synthetic and real sensor network data and obtained the operational complexity (C) vs. the distortion (D) (C-D curve) for each of the data sets⁴. In all our simulations we considered a transmission rate $R_i = 2\text{bits}$ per source and $\gamma_i = \frac{1}{N}$. The high rate quantizer partitioned the input space into 32 regions. To be fair, we used the same high rate quantizers for the proposed technique and the competitors. For all greedy-descent based design methods, we report the best performance over several random initializations (limited to 25). The maximum average complexity allowed for the decoder codebook was 1024 (10 bits)⁵.

As a competitor to our approach, we grouped the sources heuristically based on their correlations ensuring that highly correlated sources are grouped together and then independently applied conventional distributed coding within each group. We varied the number of sources in each group to obtain the distortion at different complexities. We considered three data-sets for our analysis:

1) **5 Synthetic Gaussian sources** : We first considered 5 uniformly spaced synthetic Gaussian sources, $\mathcal{N}(0, 1)$, with correlation exponentially decaying with the distance. Specifically, the correlation coefficient between sources X_i and X_j , is $\rho_{ij} = \rho^{|i-j|}$. In our simulations, we assumed $\rho = 0.95$ and a training set of length $|\mathcal{T}| = 200,000$. The results reported are on a test set, also of length 200,000.

⁴We note that the well known ‘time sharing’ argument is not applicable to the complexity-distortion trade-off curve. This is because, if we time share between two operating points, the decoder needs to maintain large enough memory required by the operating point with higher complexity. Nevertheless, we continue to connect all the operating points for clarity of presentation.

⁵Note that the smallest complexity we consider is 4, as we force the bit-subset selector to select at least the 2 bits sent by the corresponding encoders.

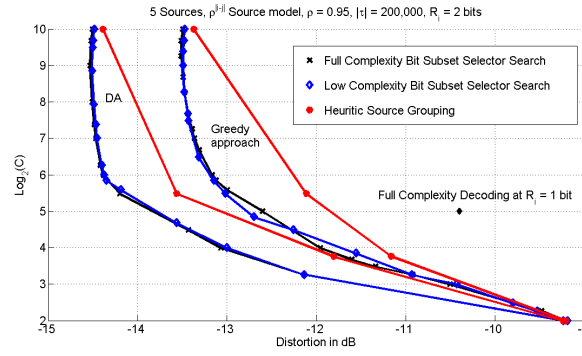


Figure 4. Complexity versus distortion trade-off for 5 Synthetic Gaussian sources

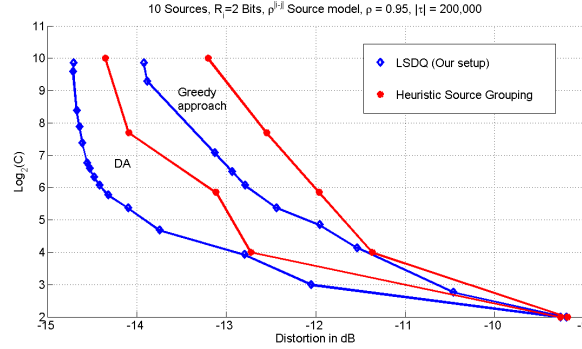


Figure 5. Complexity versus distortion trade-off for 10 Synthetic Gaussian sources

2) **10 Synthetic Gaussian sources** : As a second data set, we considered 10 uniformly spaced synthetic Gaussian sources, $\mathcal{N}(0, 1)$, again with an exponentially decaying correlation coefficient, with $\rho = 0.95$. We again chose $|\mathcal{T}| = 200,000$ for both the training and the test sets.

3) **50 Real world sensors data**: The real sensor network dataset we used was collected by the Intel Berkeley Research Lab, CA⁶. Data were collected from 54 sensors deployed in the Intel Berkeley Research Lab between February 28 and April 5, 2004. The dataset had recoding of temperature and luminescence from 25 sensors that collected the highest number of samples, which is equivalent to 50 sources. Each sensor recorded reading once every 31 s. Times when subset of these sensors failed to record data were dropped from the analysis. The data were normalized to zero mean and unit variance. Half the dataset was used for training and the remaining for testing.

The complexity-distortion trade-off curve obtained by optimizing the system using both greedy-iterative decent and using DA are shown in figures 4, 5 and 6, for all the three datasets, respectively. To be on fair grounds, we also report the results obtained by optimizing the source grouping approach using deterministic annealing. Observe that, for the 5 source synthetic dataset, it is possible to design the full complexity decoder which operates at a decoder complexity of 2^{10} . For this dataset, our approach (optimized using greedy-iterative decent) outperforms the heuristic grouping scheme by about 1 dB at a complexity of 2^7 . Further gains of close to 3 dB in distortion are

⁶The dataset is available for download at : <http://db.csail.mit.edu/labdata/labdata.html>

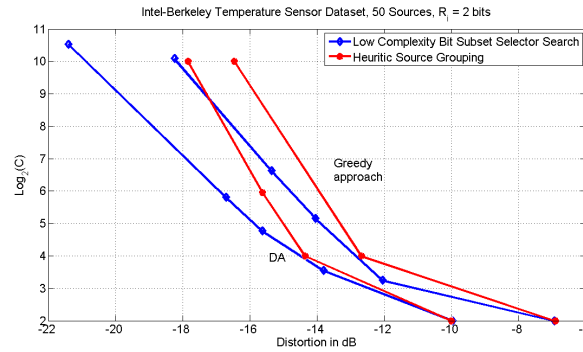


Figure 6. Complexity versus distortion trade-off for Intel-Berkeley temperature sensor dataset - 50 sources

possible using the proposed design based on DA. We also report the performance of the conventional distributed source coder transmitting at a lower rate of $R_i = 1$ bit per source. The proposed approach gains close to 2.5 dB in distortion at the same decoding complexity compared to the conventional DSC transmitting at a lower rate. This demonstrates that it is indeed beneficial to transmit more bits and allow the decoder to select a subset of the received bits for decoding, rather than to transmit fewer bits.

Similar gains are seen for the other two datasets. For the 10 source synthetic Gaussian dataset, we observe close to 1 dB gain in distortion at an average complexity of 2^6 and a further improvement of 2 dB using deterministic annealing based design of the system parameters. For the real world temperature sensor dataset, we see gains of close to 2 dB in distortion at a complexity of 2^{10} and a further improvement of 3 dB using deterministic annealing. Alternatively, we see a 16x reduction in decoder complexity at a distortion of -16 dB.

VI. DISTRIBUTED SOURCE CODING AND DISPERSIVE INFORMATION ROUTING

In this section, we consider a seemingly different application involving minimum cost communication of correlated sources across a network and show the broad applicability of the bit-subset selector module and the design methodologies we proposed in the context of large scale distributed compression. Recall the functionality of the bit-subset selector module we described in Section 3. Its primary purpose was to select a subset of the transmitted bits that provide the most reliable estimate for each source. By controlling the number of bits selected, the decoder complexity was traded for the reconstruction distortion. This intuition finds applicability in several scenarios beyond the problem of large scale distributed compression. In fact, the motivation for using the bit-subset selector module was originally derived from its applications in databases for fusion storage and selective retrieval of correlated sources [17]. Here, we focus on a seemingly different application called dispersive information routing (DIR) which involves communicating (routing) correlated sources across a network. Exactly the same underlying principles will be applicable here for efficient joint design of distributed source coders and dispersive information routers. It is important to remember that, while we demonstrate the applicability in the context of routing in sensor networks,

the underlying principles can be applied to a rich class of problems involving ‘fragmentation of information’.

Compression in multi-hop networks has gained significant importance in recent years, mainly due to its relevance in sensor network applications. Review paper [27] describes the important joint compression-routing schemes that have been developed so far. Encoding correlated sources in a network with multiple sources and sinks has conventionally been looked at from two different directions. The first approach performs compression at intermediate nodes [28], where all the information is available, without appealing to DSC principles. However, such approaches tend to be wasteful at all but the last hops of the communication path. The second approach uses distributed source coding to exploit correlation at the source nodes followed by simple routing at intermediate nodes. Well designed DSC could provide considerable performance improvement and/or complexity/energy savings. Various aspects of DSC for routing have been considered in a number of publications, and notably in [29], where the authors consider joint optimization of Slepian - Wolf coding and conventional routing.

Optimal routing schemes, designed for independent sources (conventional routing), have been studied extensively [30], primarily due to the growth of the Internet. It may be tempting to assume that an optimal distributed source code, which tries to eliminate the dependencies between sources, followed by a conventional routing mechanism, would achieve optimality in communication cost for transmitting correlated sources across a network (see e.g. [29], [31]). However, we showed recently in [32] that for a network with multiple sources and sinks, DSC followed by conventional routing suffers from inherent and significant drawbacks. We then introduced a new paradigm called “dispersive information routing” (DIR) in [20], [21], which provides significant asymptotic gains compared to conventional routing for communicating correlated sources. In this new routing paradigm, every intermediate node is allowed to split a packet and forward only a subset of received bits on each of the forward paths. In this paper, deriving principles from the problem of large scale distributed compression, we propose a practical integrated framework for distributed compression and dispersive information routing and show that the bit-subset selector module and the design methodologies we proposed in the context of large scale distributed compression, play a central role in designing such routers. We first show using a simple network example the sub-optimality of conventional routing methods and motivate the new paradigm. We then formulate the problem of joint design of DSC and DIR that allows us to use similar algorithms as in Section IV. We then apply the design to a sensor grid with multiple sources and sinks and demonstrate the potential gains of the proposed methodology. We note that unlike network coding [33], [34], DIR can be realized using conventional routers without recourse to expensive coders at intermediate nodes, making it more suitable for low powered sensor networks. We begin with the problem setup followed by a brief information theoretic motivation for the new routing paradigm. We refer to [21] for a detailed analysis on the asymptotic gains achievable using DIR.

A. Problem Setup

Let a network be represented by an undirected graph $G = (V, E)$. Each edge $e \in E$ is a network link whose communication cost depends on the edge weight w_e . The nodes V consist of N source nodes, M sinks, and $|V| - N - M$ intermediate nodes. Source node i observes X_i where (X_1, X_2, \dots, X_N) are correlated random variables. The sinks are denoted S_1, S_2, \dots, S_M . Each sink requests information from a subset of sources. Let the subset of nodes requested by sink S_j be $V_j \subseteq V$. Conversely, source i has to be reconstructed at a subset of sinks denoted $S_i \subseteq \{S_1, S_2, \dots, S_M\}$. Define traffic matrix (or “request” matrix) T , for network graph G as the $N \times M$ binary matrix that specifies which sources must be reconstructed at each sink:

$$T_{ij} = \begin{cases} 1 & \text{if } i \in V_j \\ 0 & \text{else} \end{cases} \quad (26)$$

Without loss of generality we assume that every source is requested by at least 1 sink.

The cost of communication through a link is a function of the bit rate flowing through it and the edge weight, which we will assume for simplicity to be a simple product $f(R, w_e) = R w_e$, noting that the approach is directly extendible to more complex cost functions. The objective is to design encoders, routers and decoders to minimize the overall network cost (calculated given the set of link rates and edge weights) at a prescribed level of average distortion.

We denote by E_B^i , the set of all paths from source i to the subset of sinks $B \subseteq \{S_1, S_2, \dots, S_M\}$. The optimum route from the source to these sinks is determined by a spanning tree optimization (minimum Steiner tree) [30]. More specifically, for each source node i , the optimum route is obtained by minimizing the cost over all trees rooted at node i which span all sinks $S_j \in B$. The minimum cost of transmitting a single bit from source i to the subset of sinks B , denoted by $d_i^*(B)$, is given by:

$$d_i^*(B) = \min_{P \in E_B^i} \sum_{e \in P} w_e$$

These costs will play an important role in the proposed joint DSC-DIR design in Section VI-C. However, before we describe the proposed framework, we begin with an information theoretic motivation for the new routing paradigm.

B. Information Theoretic Motivation

Fig. 7 depicts a simple network with 3 sources (X_1, X_2 and X_3) and two sinks (S_1 and S_2). S_1 requests for $\{X_1, X_2\}$ while S_2 requests for $\{X_2, X_3\}$. There is one intermediate node, c (called collector), which serves the purpose of a simple router. The sources communicate with the sinks only through the collector. This is a toy example for a large sensor network with all intermediate nodes collapsed to a single collector node. Note that we

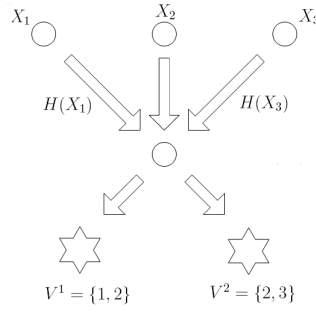


Figure 7. Example to demonstrate the asymptotic gains of DIR

motivate our approach using the loss-less setting (asymptotic behavior). But the practical design (non-asymptotic) makes it feasible to work directly within the lossy coding setting with loss-less as a special case.

The collector has to transmit enough information to S_1 so that it can decode both X_1 and X_2 and similarly enough information to S_2 to decode X_2 and X_3 . Hence the rates on the edges (c, S_1) and (c, S_2) are at least $H(X_1, X_2)$ and $H(X_2, X_3)$, respectively. Let us say the weights on the edges are such that, $w_{1,c}, w_{3,c} \ll w_{2,c}, w_{c,S_1}, w_{c,S_2}$. This implies that X_1 and X_3 transmit data at rates $H(X_1)$ and $H(X_3)$, respectively. As source X_2 has to transmit enough data for both the sinks to decode it losslessly:

$$R_2 \geq \max(H(X_2|X_1), H(X_2|X_3)) \quad (27)$$

Conventional routing methods (designed for independent sources) do not "split" a packet at an intermediate node and hence would forward all the bits from X_2 to both the sinks. This would mean sub-optimality on either one of the branches (c, S_1) or (c, S_2) if $H(X_2|X_1) \neq H(X_2|X_3)$.

But instead, let us now relax this restriction and allow the collector to route only a subset of bits on each edge. Note that instead of "splitting the packet" at the collector, we could equivalently think of source X_2 transmitting 3 smaller packets to the collector. First packet has rate $R_{2,\{1,2\}}$ bits and is destined to both the sinks. Two other packets have rates $R_{2,1}$ and $R_{2,2}$ bits and are addressed to sinks S_1 and S_2 , respectively. Technically, in this case, the collector would just have to route the received packets in a conventional manner. It can be shown using random product binning arguments (refer to [20]) that the rate tuple $(R_{2,\{1,2\}}, R_{2,1}, R_{2,2}) = (H(X_2|X_1, X_3), I(X_2, X_3|X_1), I(X_2, X_1|X_3))$ is achievable and the rates on edges (c, S_1) and (c, S_2) achieve their respective lower bounds.

We term such a routing mechanism, where each intermediate node in a multi-hop network can *route any subset* of the received bits on each of the forward paths as "dispersive information routing" (DIR). In the next section we propose a practical integrated framework for efficient design of joint distributed source coders and dispersive information routers that allows us to trade-off the average reconstruction quality against the total communication cost. Note the clear difference from network coding. DIR does not require expensive coders at intermediate nodes,

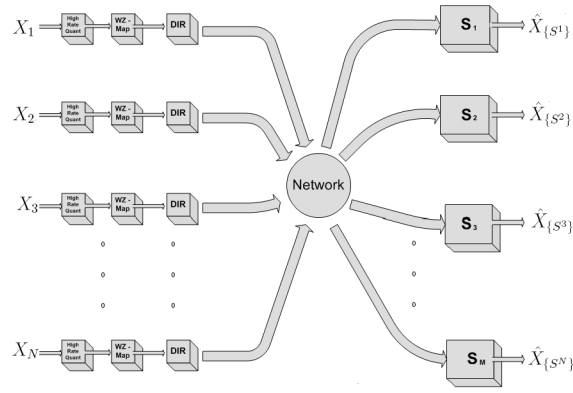


Figure 8. Integrated framework for DSC and DIR. The DIR modules decide the subset of sinks each bit must be routed to.

but rather can always be realized using conventional routers with each source transmitting multiple smaller packets into the network, intended to different subsets of sinks. Also note that such a routing mechanism is inessential when the sources are independent.

C. Integrated DSC and Dispersive Routers

Consider a network as formulated in subsection VI-A with each sink S_j requesting for a subset $V^j \subseteq V$ of sources. Recall the encoding mechanism at each source. The random variable X_i is first quantized using the high rate quantizer \mathcal{H}_i and then mapped to the transmission index I_i , using the WZ-map. Before we describe the proposed integrated framework for DSC and DIR design, we first consider two standard approaches for routing in a network and mention their drawbacks.

1) *Broadcasting*: The bits from all the sources are routed to all the sinks (broadcasted). Then the decoder \mathcal{D}_{ij} for source X_i at sink S_j can be expressed as:

$$\mathcal{D}_{ij} : \mathcal{I} \rightarrow \hat{\mathcal{X}}_{ij} \subset \mathcal{R} \quad \forall j, i \in V^j \quad (28)$$

Such a routing mechanism, though might provide good distortion performance, is highly wasteful with respect to the communication cost.

2) *Conventional Routing*: All the bits transmitted by source i are routed to its destination sinks $S_j \in S^i$. If we use $\mathcal{I}_{S_j} = \prod_{i \in S_j} \mathcal{I}_i$, to denote the set of all possible received symbols at sink S_j , then the decoder \mathcal{D}_{ij} can be expressed as:

$$\mathcal{D}_{ij} : \mathcal{I}_{S_j} \rightarrow \hat{\mathcal{X}}_{ij} \subset \mathcal{R} \quad \forall j, i \in V^j \quad (29)$$

As we saw in the previous sub-section, such a routing technique, wherein fragmentation of the packets is not allowed, is (asymptotically) suboptimal for communicating correlated sources over a network. Another major drawback with such a routing technique is that, an unrequested but correlated source may provide less expensive information about

requested sources. Hence, restricting each sink to receive packets only from the requested sources could be highly suboptimal.

Motivated by these drawbacks, the proposed dispersive information routing paradigm allows for every sink to receive information from all the sources, regardless of the request matrix. Moreover, we allow for packet splitting at intermediate nodes, i.e., a sink may receive only a subset of the bits transmitted by any given source. Essentially, *we allow each bit to be routed to any subset of the sinks*. We introduce a new module at each encoder which decides the route for each bit generated at that encoder. Note that if each bit follows the route prescribed by the encoders, every intermediate node would just be forwarding a subset of the received bits on each of the forward paths. We call this module the “dispersive information router module” to indicate the routing mechanism it induces in the network. We denote by $S = \{S_1, S_2 \dots S_M\}$ the set of all sinks and by 2^S the power set (set of all subsets) of S . Formally, the router module at the i^{th} encoder is given by:

$$\mathcal{C}_i : \{1, \dots, R_i\} \rightarrow 2^S \quad (30)$$

and denote by $\mathcal{C} = \mathcal{C}_1 \times \mathcal{C}_2 \dots \mathcal{C}_N$. The router modules uniquely determine the set of all the received bits at each sink. It is important to note the similarities between the bit-subset selector module in Section 3 and the dispersive information router modules here. In the former case, the bit-subset selector decides the subset of received bits to be used to estimate each source and the objective is to minimize the average reconstruction distortion subject to a constraint on the total decoder complexity. Whereas in the latter case, the dispersive information router modules decide the subset of the transmitted bits that are routed to each sink, to estimate the requested sources. The objective is to design the router modules (jointly with the encoders and the decoders) to minimize the average distortion, subject to a constraint on the total communication cost. The underlying principles in both these setting are exactly the same, though the explicit cost functions are different. Hence, we use the same principles described in Section IV for the joint design of encoders, routers and the reconstruction codebooks. Due to the close relation between the two problems, we only state the Lagrangian cost function to be optimized in the DIR framework and omit restating the details of the design algorithms.

The decoder at each sink is now modified to be the mapping:

$$\mathcal{D}_{ij} : \mathcal{I} \times \mathcal{C} \rightarrow \hat{\mathcal{X}}_{ij} \subset \mathcal{R} \quad \forall j, i \in V^j \quad (31)$$

The total communication cost W of the system is given by:

$$W = \sum_{i=1}^N \sum_{j=1}^{R_i} d_i^* (\mathcal{C}_i(j)) \quad (32)$$

and the average reconstruction distortion is measured as:

$$D = E \left[\sum_{j=1}^M \sum_{i=1}^{|V^j|} \gamma_{ij} d_{ij}(X_i, \hat{X}_{ij}) \right] \quad (33)$$

where $d_{ij} : R \times R \rightarrow [0, 1)$ are well-defined distortion measures and γ_{ij} are used to weigh the relative importance of each source at each sink to the total distortion. If we specialize the distortion metric to be the mean squared error and assume the expectations to be approximated by empirical averages, the average distortion is measured as,

$$D = \frac{1}{|\mathcal{T}|} \left[\sum_{\mathbf{x} \in \mathcal{T}} \sum_{j=1}^M \sum_{i=1}^{|V^j|} \gamma_{ij} (x_i - \mathcal{D}_{ij}(I, \mathcal{C}))^2 \right] \quad (34)$$

where $\mathbf{x} = \{x_1 \dots x_N\}$ and $I = [\mathcal{E}_1(x_1), \mathcal{E}_2(x_2) \dots \mathcal{E}_N(x_N)]$ denotes the set of all bits being routed in the network. The trade-off between the distortion and the communication cost is controlled using a Lagrange parameter $\lambda \geq 0$ to optimize the weighted sum of the two quantities. From (34) and (32), the Lagrangian cost to be minimized is:

$$L = D + \lambda W \quad (35)$$

where D and W are given by (34) and (32), respectively. The objective is to find \mathcal{E}'_i s, \mathcal{C}'_i s and \mathcal{D}'_{ij} s that minimize L for a given λ . We again note that this optimization framework is exactly same as that in Section III, but with a different Lagrangian cost and hence we omit restating the design algorithms. We also note that, for large networks, it would be necessary to explicitly control the decoder complexity while trading-off distortion and communication cost. This could be easily performed by introducing multiple Lagrange parameters to optimize the weighted sum of all the three quantities. We omit the details here as it is a direct extension of the above formulations. We note that the above approach for the design of the dispersive information router modules is centralized in the sense that the optimization is done offline, at a central location, before the operation. In this paper, we aim to establish the potential gains from using DIR in a practical setting, which in turn promotes future research for developing efficient decentralized design strategies.

D. Results for Dispersive Information Routing

We considered three sensor network grids to demonstrate the gains achievable during dispersive information routing compared to the conventional routing techniques. We assumed 4 sinks for all three grids, placed at the 4 corners of a square grid of length $d_0 = 100$. The first two datasets were synthetic Gaussian datasets obtained by random deployment of the sensors and intermediate nodes, to mimic a real world scenario with inaccessible regions. The correlation between two sensors at a distance d was assumed to be ρ^{d/d_0} . We assumed $\rho = 0.8$ for both these experiments. The first simulation had 4 sources with 10 other intermediate nodes, while the second simulation assumed 8 sources and 15 other intermediate nodes. For both these simulations, the training set size

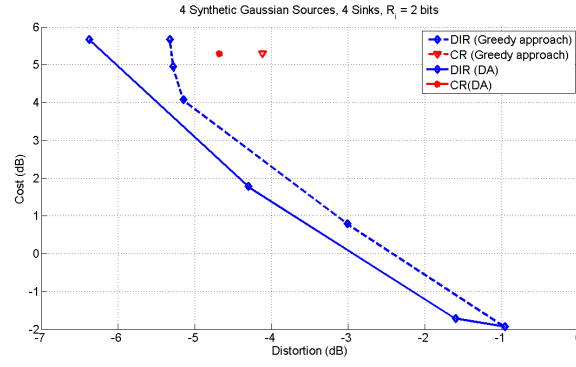


Figure 9. Cost versus distortion trade-off for 4 Randomly deployed Gaussian sources and 4 sinks

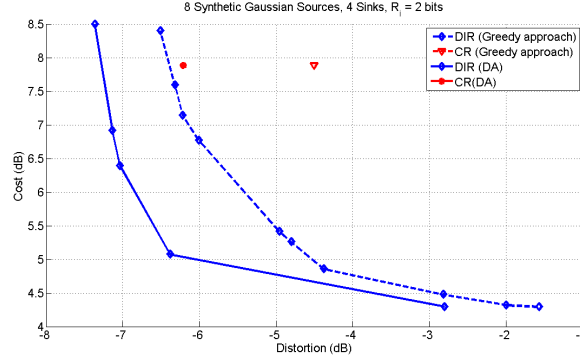


Figure 10. Cost versus distortion trade-off for 8 Randomly deployed Gaussian sources and 4 sinks

was assumed to be 200,000 samples. The third dataset we considered consisted of readings from 8 temperature sensors chosen randomly from the dataset collected by the Intel Berkeley research center. All the other 46 sensors were regarded as intermediate nodes with simple routing capabilities. The 4 sinks were assumed to be at the four corners of the building. Again, half the samples were used for training and the remaining half for testing. For a source transmitting at a rate R_i , the high rate quantizer partitioned the source space into 2^{R_i+3} regions, for e.g. if the source rate was 2 bits, $N_i = 32$. For designs based on greedy-iterative decent based methods, we report the best performance over several random initializations (limited to 25). We used the square of the distance between two

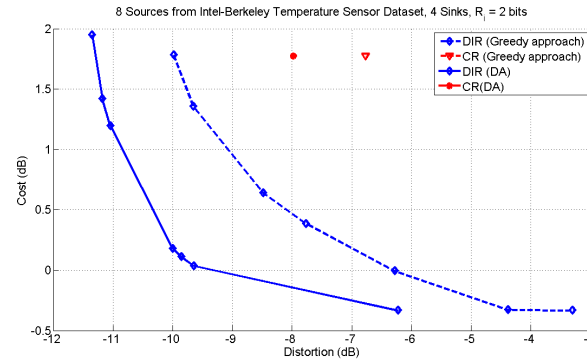


Figure 11. Cost versus distortion trade-off for Intel-Berkeley temperature sensor dataset, 8 sources and 4 sinks

sensors as the corresponding edge weight (w_e) for all the simulations. The Steiner tree optimized costs (d_i^*) were computed before the design of the modules, using an optimal Steiner tree algorithm. We note that the minimum Steiner tree problem is NP-hard and requires approximate algorithms to solve in practice for larger networks.

An operational cost-distortion trade-off curve was obtained using the proposed design techniques for all the three datasets. Figures 9, 10 and 11 show the C-D curves for the three datasets, respectively. As a competitor, we show that performance of optimally designed DSC for conventional routing. As it is clearly evident, DIR gains close to 2 dB in distortion for the 4 source and 1.5 dB in distortion for the 8 source synthetic grids, respectively. For the real sensor network dataset, we see higher gains of above 3 dB compared to the conventional routing technique. It is also evident that the DA based designs show significant reduction in communication cost at a fixed reconstruction distortion. The gains in communication cost due to DA based design for the three datasets are 1 dB, 2 dB and 1.5 dB, respectively.

VII. CONCLUSION

We proposed a new decoding paradigm for large-scale distributed coding which operates at practical codebook complexities. We formulated a Lagrangian cost and proposed a design algorithm based on deterministic annealing (DA) to optimize the performance trade-off between complexity and distortion. We then pointed out close similarities between the problem of large scale distributed coding and the problem of dispersive information routing for communicating correlated sources across a network. We used similar design principles proposed in the context of large scale distributed coding for the design of an integrated framework for distributed source coders and dispersive information routers. Simulation results on both real and synthetic datasets show considerable gains of the proposed approaches over conventional methods.

REFERENCES

- [1] D. Slepian and J. K. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. on Information Theory*, 19:471–480, Jul 1973.
- [2] A. D. Wyner and J. Ziv. The rate-distortion function for source coding with side information at the decoder. *IEEE Trans. on Information Theory*, 22:1–10, Jan 1976.
- [3] T.S. Han and K. Kobayashi. A unified achievable rate region for a general class of multiterminal source coding systems. *IEEE Trans. on Information Theory*, IT-26:277–288, May 1980.
- [4] S.S. Pradhan and K. Ramchandran. Distributed source coding using syndromes (DISCUS): Design and construction. *IEEE Trans. on Information Theory*, 49(3):626–643, Mar 2003.
- [5] Z. Xiong, A. Liveris, and S. Cheng. Distributed source coding for sensor networks. *IEEE Signal Processing Magazine*, 21(5):80–94, 2004.
- [6] J. Bajcsy and P. Mitran. Coding for the slepian-wolf problem with turbo codes. In *Proceedings of IEEE GLOBECOM*, number 2, pages 1400–1404, 2001.

- [7] T.J. Flynn and R.M. Gray. Encoding of correlated observations. *IEEE Trans. on Information Theory*, 33(6):773–787, 1987.
- [8] D. Rebollo-Monedero, R. Zhang, and B. Girod. Design of optimal quantizers for distributed source coding. In *Proceedings of IEEE Data Compression Conference (DCC)*, pages 13 – 22, Mar 2003.
- [9] M. Fleming, Q. Zhao, and M. Effros. Network vector quantization. *IEEE Trans. on Information Theory*, 50:1584–1604, Aug 2004.
- [10] A. Saxena, J. Nayak, and K. Rose. Robust distributed source coder design by deterministic annealing. *IEEE Trans. on Signal Processing*, 58(2):859–868, Feb 2010.
- [11] P. Yahampath. Predictive vector quantizer design for distributed source coding. In *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, volume 3, pages 629–632, Apr 2007.
- [12] A. Saxena and K. Rose. Distributed predictive coding for spatio-temporally correlated sources. *IEEE Trans. on Signal Processing*, 57:4066–4075, Oct 2009.
- [13] J. Barros and M. Tüchler. Scalable decoding on factor graphs - a practical solution for sensor networks. *IEEE Trans. on Communications*, 55(2):284 – 294, Feb 2006.
- [14] G. Maierbacher and J. Barros. Low-complexity coding and source-optimized clustering for large-scale sensor networks. *ACM Transactions on Sensor Networks*, 5(3), May 2009.
- [15] R. Yasaratna and P. Yahampath. Design of scalable decoders for sensor networks via Bayesian network learning. *IEEE Trans. on Communications*, 57(10):2868 – 2871, Oct 2009.
- [16] J. Nayak, S. Ramaswamy, and K. Rose. Correlated source coding for fusion storage and selective retrieval. In *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, pages 92–96, Sep 2005.
- [17] S. Ramaswamy, J. Nayak, and K. Rose. Fusion coding of correlated sources for storage and selective retrieval. *IEEE Trans. on Signal Processing*, 58(3):1722–1731, Mar 2010.
- [18] S. Ramaswamy and K. Rose. Predictive fusion coding of spatio-temporally correlated sources. In *Proceedings of IEEE Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 2509–2512, Mar 2008.
- [19] S. Ramaswamy and K. Rose. Shared descriptions fusion coding for storage and selective retrieval of correlated sources. In *Proceedings of IEEE Data Compression Conference (DCC)*, pages 262–271, Mar 2008.
- [20] K. Viswanatha, E. Akyol, and K. Rose. On optimum communication cost for joint compression and dispersive information routing. In *Proceedings of IEEE Information Theory Workshop (ITW)*, pages 1–5, Sep 2010.
- [21] K. Viswanatha, E. Akyol, and K. Rose. An achievable rate region for distributed source coding and dispersive information routing. In *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, pages 776–780, Aug 2011.
- [22] S. Ramaswamy, K. Viswanatha, A. Saxena, and K. Rose. Towards large scale distributed coding. In *IEEE Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 1326 – 1329, Mar 2010.
- [23] K. Viswanatha, E. Akyol, S. Ramaswamy, and K. Rose. Distributed source coding and dispersive information routing: An integrated approach with networking and database applications. In *Proceedings of European Signal Processing Conference (EUSIPCO)*, pages 1894–1898, Aug 2010.
- [24] K. Viswanatha, S. Ramaswamy, A. Saxena, and K. Rose. Error-resilient and complexity constrained distributed coding for large scale sensor networks. In *Proceedings of ACM/IEEE conference on Information Processing in Sensor Networks (IPSN)*, Apr 2012.
- [25] P. Yahampath. Joint source decoding in large scale sensor networks using Markov random field models. In *Proceedings of IEEE Conference on Acoustics Speech and Signal Processing (ICASSP)*, pages 2769 – 2772, Apr 2009.
- [26] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of IEEE*, 86(11):2210–2239, Nov 1998.
- [27] H. Luo, Y. Liu, and S. K. Das. Routing correlated data in wireless sensor networks: A survey. *IEEE Network*, 21(6):40–47, Dec 2007.

- [28] S. Patten, B. Krishnamachari, and R. Govindan. The impact of spatial correlation on routing with compression in wireless sensor networks. *IEEE Trans. on Sensor Networks*, 4(4), 2008.
- [29] R. Cristescu, B. Beferull-Lozano, and M. Vetterli. Networked Slepian-Wolf: Theory, algorithms and scaling laws. *IEEE Trans. on Information Theory*, 51(12):4057–4073, Dec 2005.
- [30] T. Cormen, C. Leiserson, and R. Rivest. *Introduction to algorithms*. McGraw-Hill Science/Engineering/Math, Jul 2001.
- [31] J. Liu, M. Adler, D. Towsley, and C. Zhang. On optimal communication cost for gathering correlated data through wireless sensor networks. In *Proceedings of the 12th ACM Annual International Conference on Mobile Computing and Networking*, pages 310–321, Sep 2006.
- [32] K. Viswanatha, E. Akyol, and K. Rose. Towards optimum cost in multi-hop networks with arbitrary network demands. In *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, pages 1833–1837, Jun 2010.
- [33] A. Ramamoorthy, K. Jain, P.A. Chou, and M. Effros. Separating distributed source coding from network coding. *IEEE Trans. on Information Theory*, 52(6):2785–2795, 2006.
- [34] D.S. Lun, M. Medard, T. Ho, and R. Koetter. Network coding with a cost criterion. In *Proceedings of International Symposium on Information Theory and its Applications*, pages 1232–1237, 2004.